A Comprehensive Strategy for Designing a Web-Based Medical Curriculum

Jeff Zucker, M.A.[1], Herbert Chase, M.D.[2], Pat Molholt, M.L.S.[1],
Carol Bean, M.L.S., Ph.D.[3], Robert M. Kahn, Ph.D[1].
Office of Scholarly Resources[1], Department of Medicine[2], Department of Medical Informatics[3]
Columbia University, New York, NY

In preparing for a full featured online curriculum, it is necessary to develop scaleable strategies for software design that will support the pedagogical goals of the curriculum and which will address the issues of acquisition and updating of materials, of robust content-based linking, and of integration of the online materials into other methods of learning. A complete online curriculum, as distinct from an individual computerized module, must provide dynamic updating of both content and structure and an easy pathway from the professor's notes to the finished online product. At the College of Physicians and Surgeons, we are developing such strategies including a scripted text conversion process that uses the Hypertext Markup Language (HTML) as structural markup rather than as display markup, automated linking by the use of relational databases and the Unified Medical Language System (UMLS), integration of text, images, and multimedia along with interface designs which promote multiple contexts and collaborative study.

INTRODUCTION

The medical curriculum at the College of Physicians and Surgeons, like that at many other medical schools, is shifting from large lectures to individualized learning, from the traditional discrete subject matters towards integrated courses, from memorization toward learning strategies, and from classroom-bound experiences to those oriented to clinical/social contexts beyond the classroom. First-year students previously learned basic science in lecture formats for biochemistry, physiology, histology, genetics and anatomy. The first four of these are now combined into a single year-long integrated course - Science Basic to the Practice of Medicine and Dentistry - that includes the participation of some 60 faculty members. Starting with the class of the year 2000, next year's first-year students, this course will include a required component of online web-based curriculum.

If the online curriculum is to support and enhance the pedagogical goals of the course revision, it too must aim to encourage self-paced, strategy-oriented, integrated and extra-classroom learning. The goal of the online curriculum is thus not to become an online textbook, or even an online library, but an online learning environment, similar to the University itself.

Such grandiose goals require broad and robust software design strategies. We have made a start at developing such strategies in several areas: 1) a data acquisition process which attempts to build on what faculty members already do in preparing their class notes and which is capable of handling constant revision and updating; 2) a data conversion process which places emphasis on the structure of the materials and which feeds back a structured version of the material to the faculty who produced the originals; 3) a data storage process which attempts to recognize that it is impossible to predict in advance all of the ways the information will be used; 4) a data linking and retrieval method that is based on concepts and semantic networks rather than on individual hand-coded links; 5) a presentation and user interface strategy that emphasizes context sensitivity, multiple views, multiple user needs and collaborative communication among a variety of user groups.

As we move forward to implement these strategies, HTML and the World-Wide-Web are not simply the tools at hand, rather they are central to the design strategy at several levels. The ubiquity and ease of use of these tools provides us with the ability to use them to blur the edges of authorship on two ends: on the one hand, as the faculty become more sophisticated in the use of HTML, they themselves will take a more active role in the actual software design of the system and on the other, as the students become more sophisticated with it, they will begin to create their own webs of information within and around the published curriculum. The Web will serve as the tool to create the online curriculum, the tool to view it, and the tool to critique, discuss and change it.

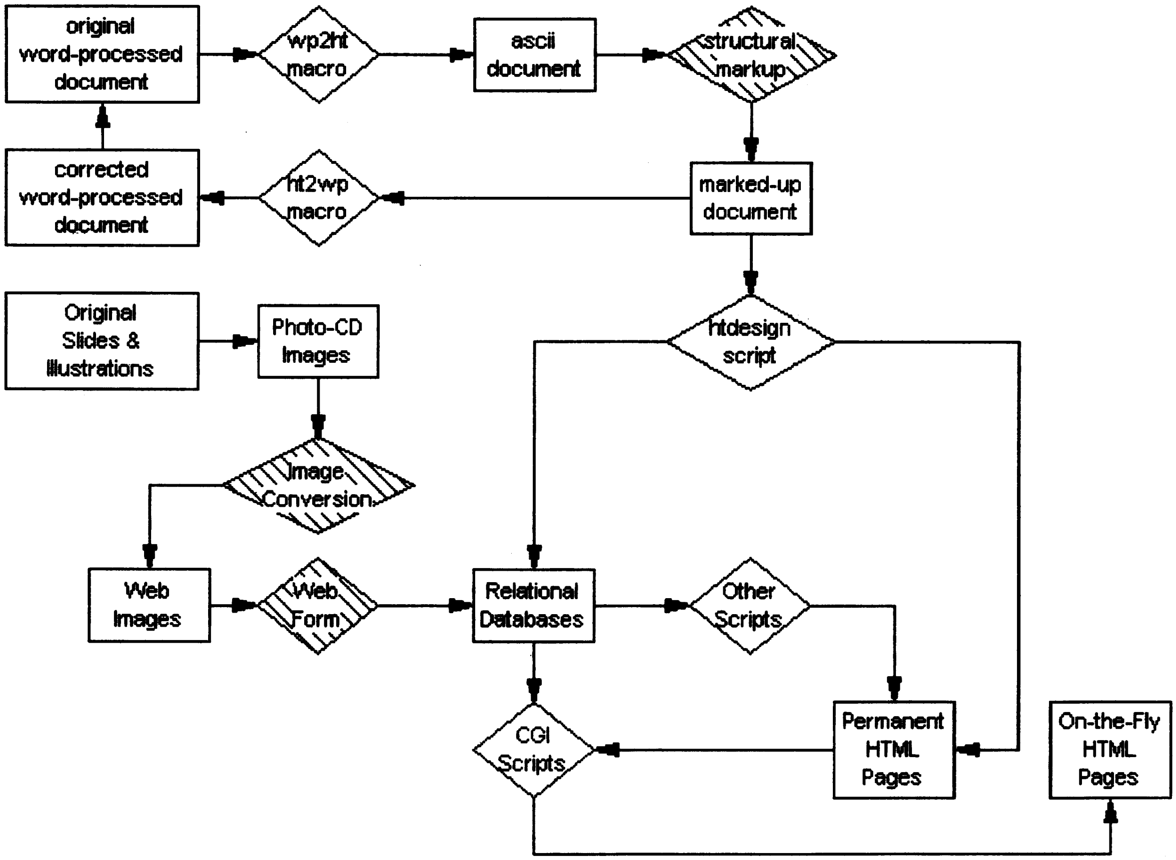

Figure 1: Curriculum acquisition, conversion, storage, retrieval, and presentation

BACKGROUND

At the Office of Scholarly Resources of the Columbia-Presbyterian Medical Center (CPMC), we have been using hypertext applications[1], Internet-based resources[2], and combined HTML/database techniques[3] to present scholarly and other information since 1992. The Department of Medical Informatics at CPMC has used the UMLS[4], to develop applications based on controlled medical vocabulary[5]. Researchers elsewhere have suggested using the UMLS and controlled vocabulary in curricular projects[6], combining static databases with dynamically built HTML pages[7], and integrating image databases into HTML pages[8]. Building on each of these approaches, we are constructing a strategy for the online curriculum which seeks to combine the strengths of the Web, of medical vocabulary and semantic network research, and relational database design.

ACQUISITION AND CONVERSION

Figure 1 is an overview of the process of acquiring, converting, storing, retrieving, and presenting the curriculum materials. The process starts with what the faculty (in most cases) already has: typed lecture notes and a slide carousel of images.

A set of word-processing macros we have designed extract the text, preserving features such as character emphasis, footnotes, formulas, etc. We then use a modified HTML to mark up the text. This markup uses HTML as a structure marking, not as a display marking. In particular the heading tags (H1...H6) are used to indicate hierarchical levels of outline structure rather than to indicate font size changes. In addition, we specify tags to indicate glossary items, learning objectives, recommended reading sections of text, and to mark out figures, tables and equations. Once we are finished with the markup, we send the page through a second word-processing macro which converts it back into its original word processing format with only the markup tags that cannot be converted into a word-processing format. This means that the faculty members get back text in a format that they can read and modify that has the added value of the structural markup we have done. In cases where the faculty member has omitted needed sections, (e.g. has not indicated a learning objective or a set of vocabulary terms), the marked up document serves as a reminder of work to do. This feedback loop in which the marked up document goes back to the faculty in word-processing format allows a richer editing process for the faculty, as well as saving time on markup when the faculty member returns the document with future revisions and updates.

The final ASCII version of the marked up document is then fed through a PERL script which places tagged sections into a series of relational databases and also creates the permanent HTML pages as they will be displayed to the user. By putting the work of developing display pages into the script we accomplish four objectives: we save time in the markup process; we gain the ability to modify the user interface and page design centrally; we keep the original document uncluttered with display features so it can be easily edited by the content providers; we gain the ability to regenerate the entire curriculum or any portion of it whenever content changes are provided. By creating the display pages at the same time as the database we ensure that all links will be correct and active - the same script that creates the display HTML pages and their locations, stores those locations along with other information about the pages and their contents in the databases thus avoiding the possibility of multiple file names and locations.

Parallel with the text acquisition and conversion process, we also convert and acquire images in a manner that feeds back a richer set of materials to the information providers, that uses the Web itself as part of the acquisition process, that allows creation of display pages to be automated in conjunction with a set of databases. Slide sets and scanned images are converted to Photo-CD format for archiving and ease of use. The images are then converted to JPEG format in three different sizes - a thumbnail sketch, a half-screen size meant to be displayed along with accompanying text, and a full screen image for detailed viewing.

Once the images are available in JPEG format, we can re-present them to the information provider in the context of a web-based form entry. A thumbnail of the image along with fields for image type, anatomical identification, and concept coding is presented so that the faculty member can identify the picture as well as provide information on the lectures and labs the image relates to, and one or more titles and captions for the image to be used in different contexts. The information from the form is stored in relational databases which are used by the faculty for sorting and selecting images and by the development team for constructing display pages. The display pages for the images, like those for the text, are constructed automatically by retrieving the image names, locations, captions, and concepts from the databases.

A third source of data for the curriculum databases is the library catalog records of multimedia resource holdings of the Columbia University libraries. The online library catalog is searched for multimedia holdings related to the health sciences and then added to the databases. This allow the online curriculum to also provide access to offline materials such as video tapes as well as to stand-alone CD-ROMs or local network resources that will be available through web browsers even if not actually on the web themselves.

STORAGE AND RETRIEVAL

Once the texts, images, and multimedia resources are acquired and converted, they are stored in relational databases. The databases include both static catalogs of all the materials and dynamic lecture and lab oriented databases. For example, a given image exists once in a static database listing its unchanging properties - its resolution, magnification, copyright status, and concept descriptors for its anatomical and functional properties. The same image might also occur in multiple rows of a second dynamic database of lectures and labs. This database stores a unique identifier to allow lookups into the static database as well as items specific to its use such as captions and titles. The image might occupy several rows in the database, each occurrence representing its use in a different lab or lecture. The separation between the two databases is another method of accommodating the dynamic nature of the curriculum. As the faculty revise their list of which images are to be used in a given lecture or lab, the dynamic database will be updated without requiring updating of the more extensive static database.

In addition to providing an easy method of updating the content of the curriculum, the relational databases also allow easy updating of the structure of the curriculum. Not only will the text of individual Web pages change, but also, the links between that page and other pages will change. Rather than hard-coding the links as part of the HTML page design, the links will be generated dynamically from the databases. Common Gateway Interface (CGI) scripts will be able to take the identifier for a given text or image on an HTML page, and use that to search the databases for links to related pages. Even if the content of that page remains constant from year to year, the links from the page can change as the database itself grows richer.

The databases will have three major structuring components: a course/lecture orientation which allows every text section, image, and multimedia module to be connected to one or more lectures in a specific course; a body systems orientation which will use an extended anatomical systems approach based on portions of the UMLS; and a wider conceptual approach including portions of the UMLS relating to both normal and pathological functioning as well as eventually to diagnostic and clinical areas of the UMLS. Details of this approach will be presented elsewhere. For the purposes of this paper, however it is important to recognize that the relational databases and the navigational linking based on them will be structured by a semantic network of terms rather than on the more limited keyword approach of traditional databases. The goal is to develop a "curriculum browser" which is capable of browsing the semantic network of the curriculum in a similar fashion to the clinically oriented browser for the UMLS as a whole.

PRESENTATION

The relational databases allow a high measure of context sensitivity within the presentation interface. A user looking at a given section of text will be presented with buttons for glossaries, multimedia options, and discussion groups. Each button will lead to items specifically related to their current context, including their current level of granularity. For example, users on the front page of the curriculum, at the top of the renal physiology section, or on the section on osmolality will be presented with three different glossary items: the first with an overall medical glossary, the second with glossary related only to renal physiology and the third with a subset of that related to osmolality. Similarly, it will be possible to design tutorials and self-learning modules outside of the lecture notes which point back into the notes themselves. For example, a vocabulary self-test module presents the user with a series of glossary items retrieved from the database of glossary terms. When the user wishes to review the terms in context, a button activates a CGI script which finds the term in the relational databases and retrieves pointers from it back into the text sections and images related to that item. This context sensitivity feature will also be used to allow faculty and students to form a series of individual and collaborative notes on the curriculum. From any text area or image, the user can press a discussion button which will allow them to take individual notes on that particular item or to query other users. This will allow faculty members to comment to each other about overlaps and inter-relationships between their areas as well as allow group discussions amongst students and student-faculty dialog. Since the medium for these dialogs will itself be the Web, the dialogs become an enriching text which sits alongside and comments on the original curriculum materials.

CONCLUSION AND FUTURE DIRECTIONS

The strategies we have outlined here are meant to cover the complete range of software development from acquisition of the original materials through to the end-user interface. By building in a dynamic, structured, concept-based approach into each phase of the project, we hope to develop a system that is flexible enough to continually grow and to be useful to a variety of users. As the project goes online this fall, we will implement evaluation procedures and gather feedback from students and faculty which will guide us with the next phases of design.

Acknowledgments

This work was supported in part by grants from the New York State Center for Advanced Technology (CAT) and the Fund for the Improvement of Post-Secondary Education (FIPSE).

References

1. Zucker J., Kahn R. and Natarajan N. Clinical, Scholarly and Campus Information Hypertext Tools At Columbia-Presbyterian Medical Center. Proceedings - Annual Symposium on Computer Applications in Medical Care 1993, 539-548.

2. Kahn RM., Molholt P. and Zucker J. CPMCnet: An Integrated Information and Internet Resource. Proceedings - Annual Symposium on Computer Applications in Medical Care 1994, 98-102.
3. Zucker J, Kahn, RM, and Molholt P. An Electronic PDR Using Fielded Searches in HTML. Proceedings - Annual Symposium on Computer Applications in Medical Care 1995, 1012.
4. Humphreys BL, Lindberg DA. The UMLS project: making the conceptual connection between users and the information they need. Bulletin of the Medical Library Association 1993 Apr;81(2):170-7
5. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical vocabulary. Journal of the Medical Informatics Association 1994, 1(1):35-50.
6. Kanter SL. Using the UMLS to represent medical curriculum content. Proceedings - the Annual Symposium on Computer Applications in Medical Care 1993, 762-5.
7. McEnery KW, Roth SM, Kelley LK, Hirsch KR, Menton DN. A method for interactive medical instruction utilizing the World Wide Web. Proceedings - the Annual Symposium on Computer Applications in Medical Care 1995, 502-507.
8. Bradley SW, Rosse C, Brinkley JF. Web-based Access to an Online Atlas of Anatomy: The Digital Anatomist Common Gateway Interface. Proceedings - the Annual Symposium on Computer Applications in Medical Care 1995, 512-516.